

Pasos para la georreferenciación automática del Catastro de Ensenada

Laura García Juan

Universidad Autónoma de Madrid

Ángel J. Álvarez Miguel

Universidad de Alcalá de Henares

Noé Fernández Sánchez

Técnico Sistemas Informáticos

El presente trabajo forma parte del proyecto SIGECAH (Sistema de Gestión de Catastros Históricos)¹, cuyo objetivo es dotar de un entorno técnico a un tipo de información de carácter textual tratada habitual y fundamentalmente de forma manual o semimanual, cuyo objetivo clave —que no único— es dotar de cartografía a un tipo de fuentes de carácter catastral y paracatastral que carece de ella, aunque la información territorial contenida en ellas constituye auténticos planos textuales por la calidad y cantidad de la misma. Asimismo, pretende convertirse en una herramienta y un referente para el estudio, gestión y análisis de

este tipo de fuentes geo-históricas. En esta primera fase de la investigación y diseño de la herramienta, el estudio se centra en una única fuente de este tipo, el Catastro de Ensenada, por cuanto, probablemente, es el mejor ejemplo de este tipo de fuentes documentales. El máximo exponente del proyecto es el portal web SIGECAH_ [www.sigecah.geo.uam.es], a través del mismo el lector interesado puede conocer más a fondo el proyecto a la par que ofrece soluciones técnicas para la introducción y análisis de los datos contenidos en la fuente. La web contiene también un servidor cartográfico (actualmente en desarrollo) que va mostrando los resultados obtenidos de aplicar los algoritmos creados en la investigación, consistentes en la simulación cartográfica, así como también las capas vectoriales de las que parten los algoritmos.

¹ El presente trabajo se ha realizado en el marco del proyecto de investigación CSO-2011-29027-C02, de la Dirección General de Investigación y Gestión del Plan Nacional de I+D+I del Ministerio de Economía y Competitividad, dirigido por la profesora C. Camarero Bullón (UAM).

Dentro de este proyecto, un hito importante ha sido la generación de una base de datos en la nube, abierta a la comunidad científica. En el número 72 de esta misma revista dimos a conocer el modelo de datos desarrollado para la gestión de la información textual del Catastro de Ensenada y presentamos ahora la extensión espacial del mismo².

La base de datos SIGECAH se desgrana en dos componentes: el «componente Ensenada», que es en el que se centra el mencionado trabajo, y el «componente GIS», que es el que nos ocupa y que se corresponde con la extensión cartográfica del anterior³. Como hemos adelantado, el objetivo final es dar una respuesta técnica y unitaria al tratamiento de la excelente información territorial contenida en el Catastro de Ensenada. Es sobradamente conocido, a pesar de la calidad de su información, que la fuente presenta un serio problema, y es que no cuenta con cartografía asociada, que habría de ser un segundo paso en su elaboración, una vez llevada a cabo la reforma fiscal y establecida la única contribución. Como dicha la reforma naufragó, la cartografía jamás llegó a levantarse. Actualmente es impensable un catastro sin cartografía, por ello nuestro objetivo es generar un sistema de gestión de catastros históricos no cartográficos, que nos permita dotar a los mismos de una cartografía, obtenida a partir de la información que contienen. Para estas fuentes se ha desarrollado una metodología que busca reconstruir y simular una cartografía para la información textual contenida

en las mismas. Básicamente nuestro método ha permitido georreferenciar la información ensenadista y, aun más, no lo ha realizado únicamente sobre información cartográfica actual sino que ha ido un paso más allá y ha recreado a nivel de pago y con una tecnología aún en desarrollo la cartografía que nunca llegó a tener la fuente y de esta forma se han dado los primeros pasos para solventar el problema de la cartografía asociada al Catastro de Ensenada.

A continuación se procede, en primer lugar, a contextualizar brevemente la fuente catastral principal de este trabajo y, en segundo, se aborda el análisis de los principales pasos de la metodología seguida para obtener una georreferenciación de la información textual del *Catastro de Ensenada*. En este recorrido se irá mostrando al lector la forma en la que la información cartográfica se almacena y se gestiona, se describe el tratamiento que se realiza con la información de las fuentes empleadas, los productos obtenidos del mismo y cualquier otro dato relevante con el objetivo de que el lector entienda el proceso seguido. El objetivo es comenzar a conceptualizar el componente GIS desde el punto de vista del mundo real sin tener en cuenta aspectos físicos y de manera independiente al SGBD (Sistema Gestor de Base de Datos). Posteriormente, el modelo relacional y el modelo entidad-relación identificarán en estos datos en entidades y relaciones, para finalmente describir la implementación en el sistema gestor de base de datos y la alimentación del componente. Para este caso emplearemos PostgreSQL, con su extensión espacial PostGIS. Se trata de un software libre, que encaja en la filosofía del proyecto basada en el empleo de este tipo de software en un intento de abaratar costes y posibilitar la expansión dentro de la comunidad científica al no requerir licencias y posibilitar la modificación del código para adecuarlo a las necesidades de cada investigación

² GARCÍA JUAN, L. *et alii* (2011): «Modelo de datos para la digitalización y gestión de fuentes catastrales geohistóricas textuales. Aplicación al Catastro de Ensenada», en *CT Catastro*, 72, págs. 73-97 [disponible en red: http://www.catastro.meh.es/esp/ct_catastro.asp]

³ El lector debe tener en cuenta que, aunque se presenta de manera separada al modelo para la introducción de los datos textuales por razones de extensión del artículo en el que se expone, va intrínsecamente unida al anterior.

1. La fuente: el Catastro de Ensenada

La denominación *Catastro de Ensenada* hace referencia a la gran averiguación de personas, bienes, rentas y cargas llevada a cabo entre 1750 y 1756 en los territorios de la Corona de Castilla. La información acopiada en dicha pesquisa habría de ser la base para una profunda reforma fiscal, cuyo objetivo era sustituir las denominadas *rentas provinciales* por una *única contribución*, universal y proporcional a la riqueza de cada uno de los contribuyentes. La denominación se aplica también a la documentación resultante de dicha averiguación. Se trata de un gran conjunto documental, formado por algo más de 80.000 gruesos libros y legajos manuscritos, que contienen una información de carácter territorial, demográfico, económico, sociológico... única en la Europa del momento, referida a las 22 provincias que formaban la Corona castellana (unos 370.000 km²), a sus algo más de 6,5 millones de personas físicas y a un número indeterminado, pero elevadísimo, de personas jurídicas (concejos, monasterios, catedrales, hospicios y hospitales, colegios, capellanías, mayorazgos, beneficios, etc.), en ellas radicadas.

La información que contiene el *Catastro de Ensenada* se organiza en tres niveles: individual, local y provincial y en diversos documentos (*Respuestas generales, memoriales, Libros de lo real, Libros de los cabezas de casa, Estados locales y provinciales*, etc.). Los documentos más interesantes para nuestro proyecto son los *Libros de lo real* y los *memoriales*, pues en ellos ha quedado recogida buena parte de la información territorial con la que trabajamos. Todo el conjunto documental ensenadista se custodia en distintos archivos de diverso tipo: nacionales (Archivo Histórico Nacional, Archivo General de Simancas, Archivo del Reino de Galicia), archivos históricos provinciales y

municipales. Además, no existe un inventario o una base de datos central que recoja de forma sistemática y segura todos los archivos en los que se guarda documentación catastral ensenadista. Se ha tratado de dar solución a este problema modelando toda la información en el «componente Ensenada» al que hacíamos referencia al inicio y que hemos tratado en profundidad en el trabajo incluido en el número 72 de esta revista⁴.

Si bien, como hemos adelantado, la información contenida en el Catastro de Ensenada es ingente y de un alto nivel de fiabilidad y exactitud, sin embargo, en este trabajo nos centraremos en aquello de lo que carece esta fuente: la cartografía. Al margen de pequeños mapas y planos no técnicos o meros bosquejos de los términos y de las parcelas, *como aparecían a la vista*, no se cuenta con una cartografía técnica levantada en el momento de elaboración del Catastro. Ante la falta de personal cualificado para llevar a cabo dicha tarea, se decidió dejar para una segunda fase el levantamiento cartográfico. Al no llevarse a cabo la reforma fiscal prevista, tampoco se llevo a cabo el levantamiento programado⁵. Por ello, dotar a nuestro siste-

⁴ Pueden verse diversos ejemplos de asientos con información territorial, así como una breve síntesis del contenido de los distintos documentos catastrales en GARCÍA JUAN, L.; ÁLVAREZ MIGUEL, A. J.; CAMARERO BULLÓN, C., y ESCALONA MONGE, J. (2012): «Generación de una metodología para la gestión y recreación cartográfica a partir de información del Catastro de Ensenada», en *Geofocus*, 12 [en red: <http://www.geo-focus.org/>]. Una sistematización de los documentos catastrales está recogida en CAMARERO BULLÓN, C. y DURÁN BOO, I. (2002): *El Catastro de Ensenada: magna averiguación fiscal de alivios de los vasallos y mejor conocimiento de los reinos*. Madrid, Dirección General del Catastro, pág. 194 [disponible en red: <http://www.eurocadastre.org/pdf/camarero5esp.pdf>]. Una aproximación a los distintos tipos de archivos en que se custodia la documentación del Catastro de Ensenada según tipos, niveles de información y territorios a los que se refiere, puede verse en CAMARERO BULLÓN, C. (2002): «Averiguarlo todo de todos: el Catastro de Ensenada», en *Estudios Geográficos*, 248-249, págs. 529 y ss.

⁵ Sobre la cartografía del Catastro de Ensenada, véase: CAMARERO BULLÓN, C. (1998): «La cartografía del Catastro de Ensenada, 1750-1756», en *Estudios Geo-*

ma de un método de georeferenciación de la documentación constituirá un valor añadido al análisis de la información textual. En este proceso de georeferenciación distinguimos dos niveles de geolocalización: un primer nivel en el que sólo se requiere adscribir la información textual a un nivel administrativo (comunidad autónoma, provincia y localidad); se trata de una serie de datos actuales que pueden conseguirse de manera gratuita y *online*. Esta aproximación sólo permite geolocalizar la información en un punto o un lugar en el espacio de acuerdo a patrones actuales. Una vez alcanzado este objetivo, tratamos de ir más allá, y nos planteamos el innegable avance en la investigación que supondría contar con una cartografía a nivel de pago o barrio asociada a esta fuente, no sólo para ser utilizado como espacio en el que representar la información que contiene sino como base para otro tipo de investigaciones. Metodológicamente se diseñó un sistema que simulara cartográficamente aquello que la fuente describe textualmente. El procedimiento está aún en fase de desarrollo y se basa en un estudio multicriterio comparado de distintas series cartográficas, cuyo punto de origen es la cartografía del catastro actual (siglo XXI) y el final la propia fuente. Planteamos una serie de elementos y datos analizables con una escala de ponderación establecida que evalúa los cambios y continuidades que cronológicamente vamos encontrando. Aquellos elementos que obtengan un grado de incertidumbre más bajo los consideramos potencialmente compatibles con la información textual del Catastro de Ensenada y, tras proceder a enlazarlos con la fuente, obtendremos un mapa a nivel de pago para el ámbito rural o de barrio en el espacio urbano. Como hemos señala-

do, esta fase del proyecto se encuentra en desarrollo y una de las líneas de investigación futuras pasa por generar un autómata celular que realice el proceso. La creación del autómata es el resultado directo de las experiencias y datos obtenidos mediante el proceso descrito.

2. Análisis previo

El objetivo principal del *componente GIS* es permitir geolocalizar los elementos almacenados en el *componente Ensenada* y se corresponde con una de las novedades de la investigación en la que se enmarca este trabajo, la georeferenciación de la información textual del Catastro de Ensenada. Para alcanzar este objetivo, se ha diseñado una metodología, descrita en un trabajo anterior, cuyos puntos más relevantes se desgranar a continuación⁶. En la localización de la información ensenadista dentro de un mapa, debemos distinguir distintos niveles de precisión. De menor a mayor, tenemos:

- Provincia, municipio.
- Barrio o pago.
- Nivel de bien.

Veamos, a partir de la descripción tipo de una parcela incluida en el *Libro de lo real* del Catastro de Ensenada de Azuqueca de Henares (Guadalajara en el siglo XVIII, Madrid hoy), los distintos niveles (cuadro 1):

Otra tierra de seis fanegas de marco real, en el Camino de la Barca, distante de esta población como doscientos pasos, que se siembra con año de intermisión; linda por arriba con dicho camino; por avajo, con tierra de don Gerónimo Ramírez, presbítero de

gráficos, 231, págs. 245-283, y CAMARERO BULLÓN, C. (2007): «La cartografía de los catastros españoles del siglo XVIII», en VV.AA.: *La cartografía cadastral a España (siglos XVIII-XX)*. Barcelona, Institut Cartogràfic de Catalunya, págs. 39-52.

⁶ GARCÍA JUAN, L.; ÁLVAREZ MIGUEL, A. J.; CAMARERO BULLÓN, C., y ESCALONA MONGE, J. (2012): «Generación de una metodología...», *op. cit.* [en red: <http://www.geo-focus.org/>].

Mondéjar, que labra Juan de Moratilla; por el solano, con otra de don Isidro Valenzuela, vezino de Madrid, que labra Juan Alonso, vezino de esta villa;

y por serrano, con tierra de don Juan Francisco del Río, vecino de Madrid... [libro de lo real de legos, Archivo Municipal de Azuqueca de Henares].

Cuadro 1

Nivel	Temático	Espacial
Provincia	Guadalajara	Coordenadas
Municipio	Azuqueca de Henares	Coordenadas
Pago	Camino de la Barca	Coordenadas
Nivel del bien	<i>Linda por arriba con dicho camino; por abajo, con tierra de don Gerónimo Ramírez, presbítero de Mondéjar, que labra Juan de Moratilla; por el solano, con otra de don Isidro Valenzuela, vezino de Madrid, que labra Juan Alonso, vezino de esta villa; y por serrano, con tierra de don Juan Francisco del Río, vecino de Madrid</i>	Coordenadas



Con este modelo perseguimos localizar tanto un sujeto como un bien en una provincia, por ejemplo Guadalajara, en un municipio, Azuqueca de Henares y en un pago, como por ejemplo el del Camino de la Barca, y, en último lugar, ubicar el bien dentro del pago. En el Catastro de Ensenada, como en el actual, la unidad territorial de base es el «pueblo», que a veces corresponde claramente con un término municipal actual, en otras no tanto, pues bajo ese concepto se hacía referencia al término con alcabalatorio y tasmía propios. Por ello, en muchos casos aparecen propietarios de bienes en término catastrado, pero residentes en otro distinto, por lo que muy frecuentemente aparecen sujetos catastrales asociados a un municipio por su lugar de vecindad, que poseen bienes asociados a otro diferente, aquel en el que radican los bienes.

A la hora de abordar la geolocalización de los elementos, nos encontraremos con dos niveles:

- Original: en la *base de datos Ensenada* contábamos con elementos que permiten una geolocalización directa de los elementos respecto a límites administrativos actuales: municipio y provincia. El objetivo es conectar éstos con su extensión geográfica, lo que va a permitir transformarlos en un punto en el mapa. En este nivel también se incluye la cartografía histórica empleada como fuente secundaria y que, si bien como ya pudimos ver, necesita un proceso de trabajo para ser introducida en un sistema de información geográfica, éste no afecta a la información que contienen⁷.

⁷ Véase *Ibidem*.

- **Producida:** se corresponde con los productos obtenidos como resultado de aplicar los procedimientos de trabajo descritos sucintamente *ut supra* y que pueden verse en detalle en otro trabajo anterior⁸. Es uno de los objetivos principales del proyecto y trabajo que estamos desarrollando: dar respuesta cartográfica a la información textual del Catastro de Ensenada.

2.1. Original

En este grupo se engloban los productos cartográficos obtenidos tras la georreferenciación de la documentación cartográfica recogida en el archivo. Una de las novedades del sistema creado es el hecho de que no nos detenemos en contar con un ráster sino que queremos dar un paso más del simple almacenamiento de una imagen, para lo cual realizaremos una vectorización de cada una de las cartografías históricas empleadas como puntos intermedios en el proceso de recreación cartográfica. En este proceso, y por la naturaleza de los datos, la georreferenciación constituye la primera fase en el ingreso de los datos históricos en nuestro sistema de información geográfica. La capa resultante sería un raster que, como datos más significativos, debe almacenar la fuente georreferenciada, la localidad representada, fecha de elaboración de la fuente, fecha de georreferenciación, persona que lo realiza y el error medio cuadrático (RMS) resultante. La segunda fase se corresponde ya con el proceso de vectorización. Al llevar a cabo la vectorización de los elementos del mapa, éstos se almacenarán por separado teniendo en cuenta las diferentes posibilidades de representar un dato geográfico: puntos, líneas y polígonos. Concretamente los elementos que queremos resaltar son:

- **Puntos.** La toponimia es uno de los puntos más importantes de nuestro proyecto y tiene un papel protagonista en nuestro objetivo de geolocalizar precisamente topónimos y otros elementos del Catastro de Ensenada. Junto a las coordenadas que posicionan el topónimo, deben recogerse otros datos fundamentales: fuente de la que procede, designación, escala, datos de quién lo realizó, subtipo: principal serían los referentes a la toponimia mayor y secundaria (no se debe entender como menos importante, sino como un segundo nivel), la microtoponimia. Este segundo objetivo no es el principal, sino que los micro topónimos se utilizarán de ajuste (una premisa que pudimos constatar en la fase de prueba fue que a mayor cantidad de datos, mayor precisión).
- **Líneas.** En este nivel recogemos caminos y calles, de las que hay que indicar como campos asociados: fuente de procedencia, topónimo.
- **Polígonos.** Manzanas y pagos dependiendo de que se trabaje con un área urbana o rural. Los atributos que deben recopilarse coinciden con los del resto de capas: fuente de la que procede, escala.

Esta categoría permite una primera aproximación al objetivo final de georreferenciar el Catastro de Ensenada, concretamente se produciría a nivel administrativo: provincia y municipio. El modelo de datos temático (*componente Ensenada*) permite establecer una relación con la provincia y municipio histórico.

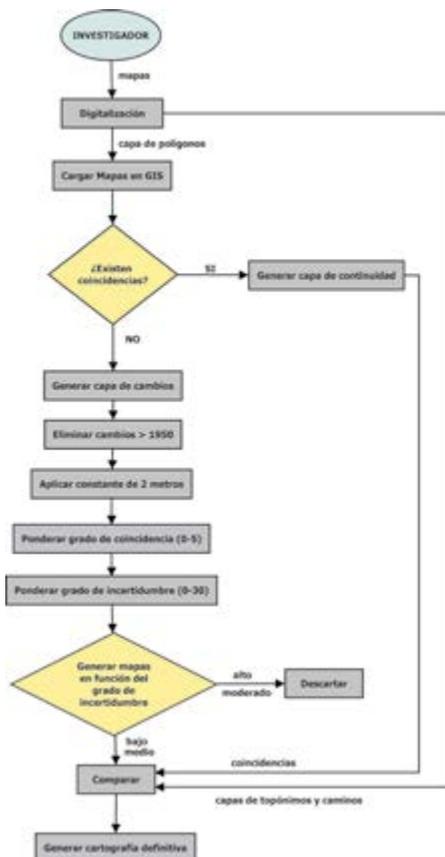
2.2. Producida

Para llegar a nivel de pago o de bien, es necesario trabajar con las capas originales, las manipularemos hasta obtener una simu-

⁸ *Ibidem.*

lación cartográfica que se asocie a la información textual del Catastro de Ensenada. Los productos derivados de este proceso son los que englobamos en esta categoría. La aplicación del procedimiento y la metodología establecidos da como resultado una capa de polígonos correspondiente a los diferentes pagos o manzanas, una de líneas que recoge las calles y caminos y, por último, una capa de puntos con los topónimos y los enclaves de interés. Asociada a la información cartográfica debemos recoger otro tipo de información como algoritmo aplicado, equipo, permisos, grado de certidumbre, denominación... (figura 1).

Figura 1



2.3. Subida de datos al servidor SIGECAH

Exceptuando las capas que hemos denominado *administrativas* (comunidades autónomas, provincias, municipios actuales y provincias históricas) el resto de datos geográficos serán producidos por cada investigador, pudiendo ocurrir que más de uno o uno mismo trabaje la misma zona con fuentes auxiliares diferentes o bien aplicando metodologías diferentes. Para que el sistema resulte cómodo, se ofrece al investigador diferentes formas de subida de datos:

- *Fichero shape*. Si el usuario no es un «poweruser» y sólo conoce ficheros más clásicos, como shape, se le permite subirlo con esta extensión.
- *Fichero SQL*. Este formato permite importar (y exportar) los trabajos realizados por uno o varios usuarios para después reintegrarlos.

El uso de formatos SQL dota al sistema SIGECAH de la posibilidad de usar herramientas ETL⁹ para el tratamiento adicional de los datos geográficos.

Para poder aceptar la subida de datos al servidor, en todos los casos se deben cumplir los criterios de normalización propuestos y los datos deben tener asociado su fichero de metadatos. Estos criterios se aplican de forma automática como parte de las facilidades que proveen los componentes PostgreSQL/PostGIS.

2.4. Metadatos

Hablar de metadatos es hablar de calidad. Los metadatos tienen por objetivo

⁹ Las siglas corresponden a las tareas: extraer, transformar y cargar. Permiten mover los datos a otros sistemas para proceder a un análisis.

permitir rastrear su origen, fuente a la que hacen referencia, autor de los mismos y sistema de proyección. En un proyecto como el que se ha descrito, que está pensado para que trabajen distintos grupos de investigación con diversas personas, se hace imprescindible tener información sobre los datos. En el caso que nos ocupa, al no existir una norma para la temática del área que nos ocupa, aplicaremos la normativa ISO 19115:2003 «*Geographic Information-Metadata*», perfil NEM (Núcleo Español de Metadatos). El núcleo español de metadatos propone un perfil abierto, conforme con la directiva INSPIRE y la ISO 19115, lo que permite una interoperabilidad de los datos tanto en el territorio español como en Europa¹⁰.

El NEM propone un perfil mínimo no restrictivo; para su utilización, cada usuario debe modelarlo para que encaje con los datos de su proyecto. Para nuestro proyecto, se ha diseñado un perfil SIGECAH, que se centra en informar sobre la persona que los ha realizado, quién los va a distribuir, en qué formato, versión y estándar del perfil de metadatos, idioma en el que se han creado, aspectos generales (escala, fecha de creación, palabras clave, resumen, sistema de referencia) y calidad (obligatoria).

3. Modelo entidad-relación

Para el tratamiento cartográfico ejecutado a través del empleo de un sistema de información geográfica, se utilizará el clásico modelo GIS vectorial. El modelo vectorial es una estructura de datos utilizada para almacenar datos geográficos. Los datos vectoriales constan de líneas o arcos, definidos por sus puntos de inicio y fin, y puntos donde se cruzan varios arcos, los

nodos, almacenados de forma explícita. Las entidades quedan definidas por sus límites solamente y los segmentos curvos se representan como una serie de arcos conectados. El almacenamiento de los vectores implica el almacenamiento explícito de la topología; sin embargo, sólo almacena aquellos puntos que definen las entidades y todo el espacio fuera de éstas no está considerado.

Hay dos formas de organizar esta base de datos doble (espacial y temática). Normalmente, los sistemas vectoriales tienen dos componentes: uno que almacena los datos espaciales y otro, los datos temáticos. A éste se le denomina *sistema de organización híbrido*, por unir una base de datos relacional para los aspectos temáticos, textuales, con una base de datos topológica, para los geográficos. Un elemento clave en este tipo de sistemas es el «*identificador*» de cada objeto o relación. Éste es único y diferente para cada objeto y permite la conexión entre ambos esquemas de datos.

Vamos a construir el modelo *Entidad-Relación* del *componente GIS* (vectorial) y el enlace con el *componente Ensenada* (temático). Como hemos adelantado, el objetivo principal del *componente GIS* es georreferenciar cada elemento del *componente Ensenada*, para lo que crearemos uno o varios elementos geográficos asociados (puntos, líneas o polígonos) en función de cada caso. Vamos a construir el modelo sin entrar en el detalle de los atributos, que explicaremos en el modelo relacional.

3.1. Identificación de entidades

En el modelo identificamos las siguientes entidades:

- *Elemento Ensenada*: cualquier elemento del *componente Ensenada* susceptible de ser georreferenciado (bienes, ubicaciones, topónimos, localidades y provincias actuales e his-

¹⁰ <http://blog-idee.blogspot.com.es/>: en esta página puede verse más información referida a bibliografía, directrices, etc. de metadatos.

- tóricas). Dichos elementos tienen un identificador único que permite localizarlos unívocamente en el modelo.
- *Elemento geográfico*: cada punto, línea o polígono georreferenciado que representa algún elemento en el *componente GIS*.
 - *Capa*: agrupación de elementos geográficos en un mismo contexto.
 - *Datos de gestión*: información sobre cada capa (autor, fuente, permisos, etc...).
 - *Metadatos*: información sobre la capa. En un primer nivel consideramos metadatos como una entidad única para simplificar el diagrama. Posteriormente detallaremos el *modelo entidad-relación* de los propios metadatos.
- *tos geográficos* (cardinalidad 1,n). La cardinalidad de la relación es también 1:N.
 - *Detallada*: una *capa* está *detallada* con unos *datos de gestión* (cardinalidad 1,1), mientras que cada ocurrencia de *datos de gestión* *detalla* una sola *capa* (cardinalidad 1,1). Ésta es una relación con cardinalidad 1:1.
 - *Descrita*: cada *capa* está *descrita* con una ocurrencia de la entidad *metadatos* (cardinalidad 1:1), mientras que cada ocurrencia de *metadatos* *describe* una sola *capa* (cardinalidad 1,1). Se trata de otra relación con cardinalidad 1:1.

A partir de la identificación de entidades y relaciones construimos nuestro modelo, recogido en la figura 2.

3.2. Identificación de relaciones

Identificamos las siguientes relaciones entre las entidades de nuestro modelo:

- *Georreferencia*: cada *elemento geográfico georreferencia* un solo *Elemento Ensenada* (cardinalidad 1,1), mientras que un *Elemento Ensenada* puede ser *georreferenciado* por uno o varios *Elementos geográficos* (cardinalidad 1, n). La cardinalidad de la relación es por tanto 1:N. Esto es debido a que un elemento del *componente Ensenada*, por ejemplo, una población, puede representarse con un elemento punto en escalas muy grandes y con un polígono en escalas menores (1,n), mientras que cada punto o polígono sólo corresponderá a un elemento concreto (1,1).
- *Situado*: cada *Elemento geográfico* se encuentra *situado* en una *capa* y sólo en una (cardinalidad 1,1), mientras que en una *capa* pueden *situarse* varios *Elementos*

3.3. Descripción del modelo entidad-relación de los metadatos

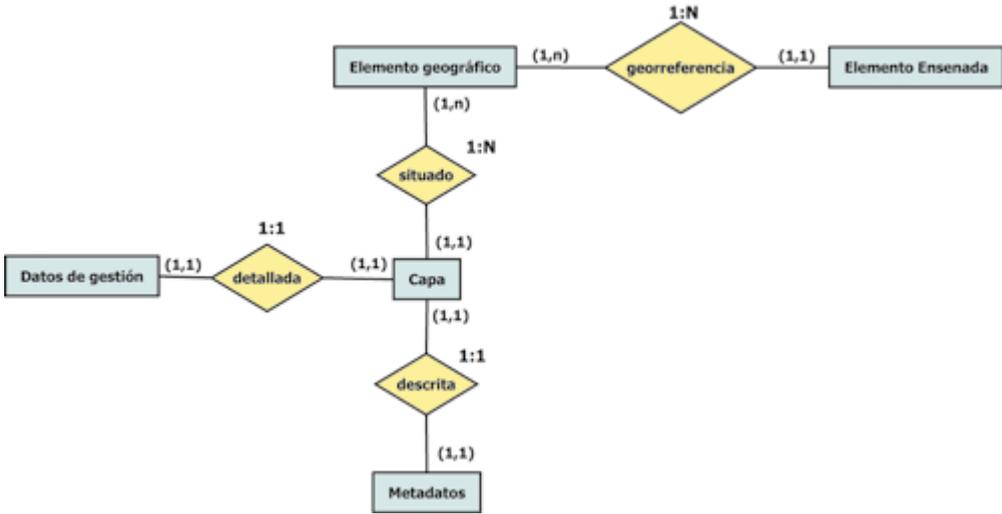
Para una mejor comprensión y un mayor detalle, se ha descompuesto de manera separada la entidad metadatos. Se ha incluido la descripción de la entidad *capas* y de la relación descrita para ilustrar el nexo de unión con el diagrama padre.

3.3.1. Identificación de entidades

En el modelo identificamos las siguientes entidades:

- *Capa*: agrupación de elementos geográficos en un contexto común determinado por la fuente y el período cronológico.
- *Metadatos*: información sobre la capa. En un primer nivel consideramos metadatos como una entidad única para simplificar el diagrama.

Figura 2



- *Calidad*: información sobre la calidad de los metadatos.
- *General*: datos globales sobre los metadatos.
- *Contacto*: información sobre el autor.
- *Distribución*: información sobre la persona encargada de expandir los metadatos.

- *Realizados*: los metadatos están *realizados* por un contacto. Se trata de otra relación con cardinalidad 1:1.
- *Distribuidos*: los metadatos están *distribuidos* por *distribución* (cardinalidad 1,1).

4. Modelo relacional

En este paso vamos a transformar el *Modelo Entidad-Relación* (MER en adelante) en un modelo relacional.

3.3.2. Identificación de relaciones

Identificamos las siguientes relaciones entre las entidades de nuestro modelo (figura 3):

- *Descrita*: cada *metadato describe* una sola *capa* (cardinalidad 1,1).
- *Definidos*: cada *metadato se encuentra definida* con una información *general* (cardinalidad 1,1).
- *Valorados*: los metadatos están *valorados* con una *calidad* (cardinalidad 1,1).

4.1. Transformación de Entidades

- La entidad *Elemento Enseñada* se transforma como una tabla o relación: *elemento_enseñada*, en la que recogemos los elementos que queremos geolocalizar del *componente Enseñada* y su clave. Esta tabla cuenta

con los atributos `id`, `tipo_elemento`, `id_elemento`)

`elemento_ensenada` (`id`, `tipo_elemento`, `id_elemento`)

- La entidad *Elemento Geográfico* se transforma en otra relación, `elemento_geografico` con los atributos `id`, `tipo_elemento`, `elemento_id`. Esta recoge los puntos, líneas y polígonos cargados en el sistema.

`elemento_geografico` (`id`, `tipo_elemento`, `elemento_id`)

- La entidad *capa* la transformamos en la tabla `capas`.

`capa` (`id`, `nombre`)

- La entidad *datos de gestión* la transformamos en la tabla `datos_gestion` con los siguientes atributos: `id`, `capa` que describe, `fuente` de los datos, `equipo` que lo introduce, `permisos` y `algoritmo` utilizado, si procede.

`datos_gestion` (`id`, `fuente`, `equipo`, `algoritmo`, `permisos`, `certidumbre`, `id_tipo_capa`, `rms`)

Como vemos, existen dependencias funcionales entre los atributos `fuente`, `equipo` y `algoritmo`, que nos indican la necesidad de normalizar esta tabla. Para ello crearemos varias tablas nuevas y propagaremos sus claves a la tabla `Datos_gestion`:

`datos_gestion` (`id`, `fuentes_id`, `equipo_id`, `id_algoritmo`, `permisos`, `certidumbre`, `tipo_capa_id`, `rms`)

`algoritmo` (`id`, `algoritmo`)

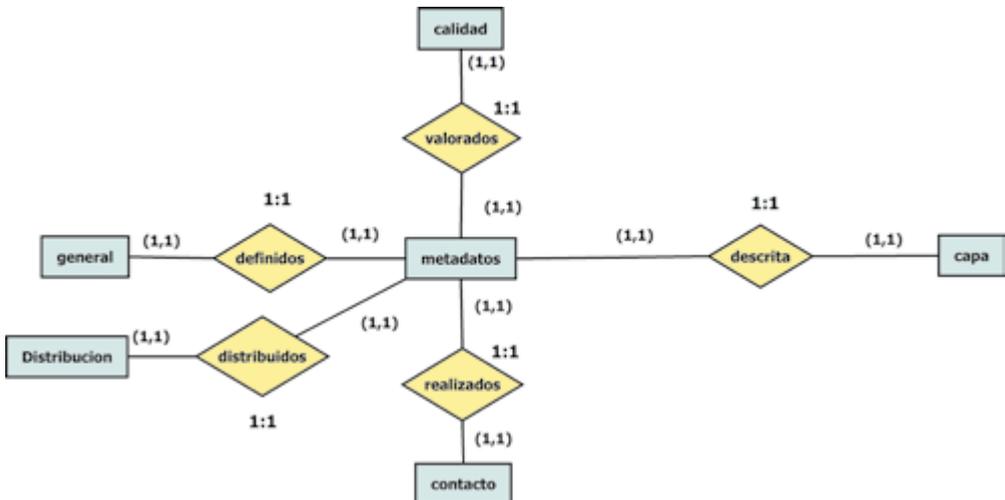
`equipo` (`id`, `equipo`)

`fuentes` (`id`, `fuentes`)

`tipo_capa` (`id`, `tipo_capa`)

- **Metadatos:** se describen a continuación como un anexo.

Figura 3



4.2. Transformación de relaciones

- Las relaciones 1:1 se transforman propagando la clave de una de las entidades en la otra. En nuestro modelo transformaremos la relación *detalla* propagando la clave de la tabla *capa* en la tabla *datos_gestion* y la relación *descrita* será transformada propagando la clave de la misma tabla *capa* a la tabla *metadatos*, con lo que las tablas *datos_gestion* y *metadatos* quedarán como sigue:

datos_gestion (id, *capa_id*, *fuente_id*, *equipo_id*, *id_algoritmo*, *permisos*, *certidumbre*, *id_tipo*, *rms*)
metadatos (id, *capa_id*...)

- Las relaciones 1:N se pueden transformar de dos formas diferentes. Vamos a utilizar ambas en función de nuestras necesidades:

- La relación *situado* será transformada propagando la clave de la entidad con una ocurrencia a la entidad con *n* ocurrencias: propagaremos la clave primaria de *capa* sobre la tabla *elemento_geografico*:

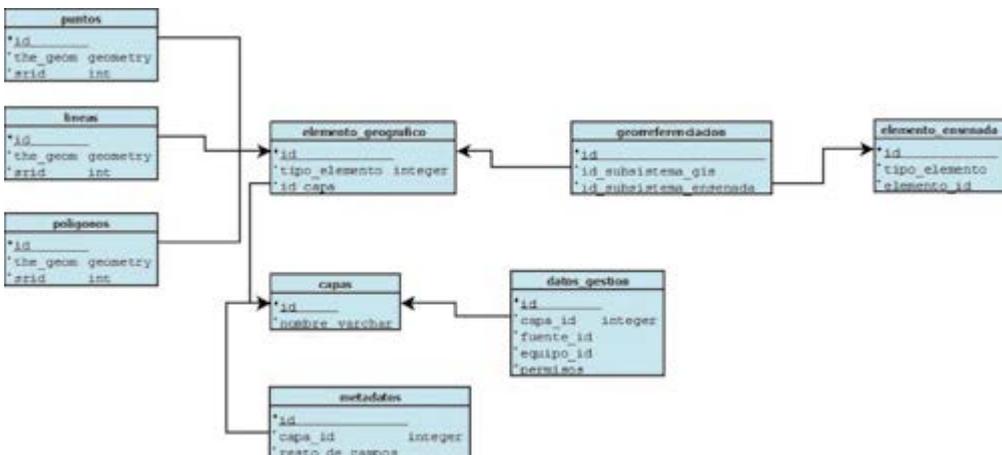
elemento_geografico (id, *capa_id*, *tipo_elemento*, *elemento_id*)

- La relación *georreferencia* hemos elegido transformarla en una tabla independiente, llamada *enlace*, a la que se propagan las claves de las entidades que relaciona:

georreferenciacion (id, *elemento_geografico_id*, *elemento_en senada_id*)

Con estas transformaciones, el modelo generado es el recogido en la figura 4.

Figura 4



4.3. Metadatos

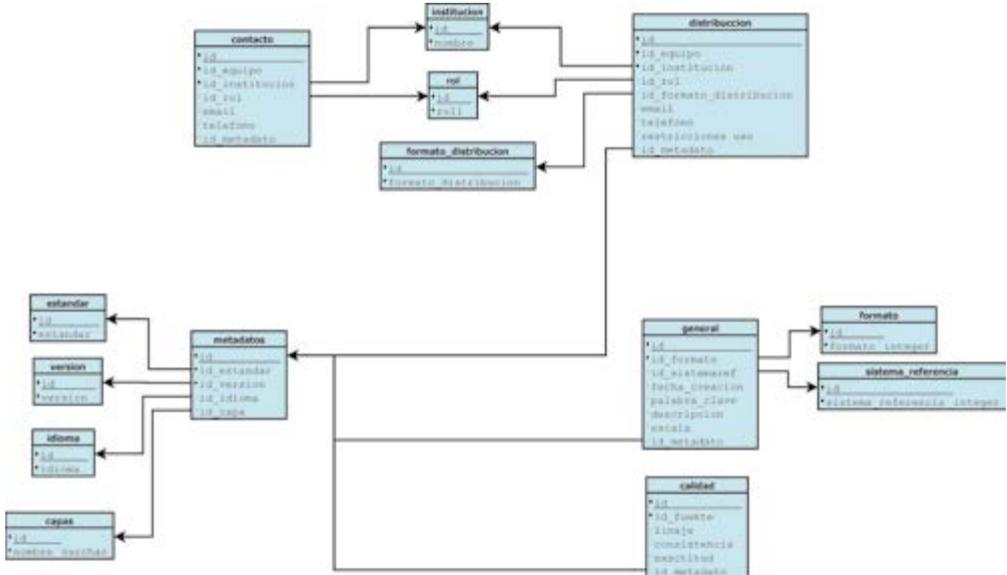
Como hicimos con el modelo entidad-relación, por su especial papel se ha desarrollado de manera separada el modelo relacional correspondiente a los metadatos.

En este paso vamos a transformar el *Modelo Entidad-Relación* (MER en adelante) en un modelo relacional, figura 5.

4.3.1. *Transformación de Entidades*

- La entidad *capa* la transformamos en la tabla **capas** con los siguientes atributos:
 - capa (**id**, nombre)
- La entidad *metadatos* la transformamos en la tabla **metadatos** con los siguientes atributos:
 - metadatos (**id**, id_estandar, id_version, id_idioma)
 - estandar (**id**, estandar)
 - version (**id**, version)
 - idioma (**id**, idioma)
- La entidad *calidad* la transformamos en la tabla **calidad** con los siguientes atributos:
 - calidad (**id**, id_fuente, linaje, consistencia, exactitud)
- La entidad *general* la transformamos en la tabla **general** con los siguientes atributos:
 - general (**id**, id_formato, id_sistema_ref, fecha_creacion, palabra_clave, descripción, escala)
 - formato (**id**, formato)
 - sistema_referencia (**id**, sistema_referencia)

Figura 5



- La entidad *contacto* la transformamos en la tabla *contacto* con los siguientes atributos:

contacto (id, id_equipo, id_institucion, id_rol, email, telefono)
institucion (id, nombre)
rol (id, rol)

- La entidad *distribución* la transformamos en la tabla *distribución* con los siguientes atributos:

distribucion (id, id_equipo, id_institucion, id_rol, id_formato_distribucion, email, telefono, restricciones_uso)
institucion (id, nombre)
rol (id, rol)
formato_distribucion (id, formato_distribucion)

4.3.2. *Transformación de relaciones*

- Las relaciones 1:1 se transforman propagando la clave de una de las entidades en la otra. En nuestro modelo transformaremos la relación *descrita* propagando la clave de la tabla *capa* en la tabla *metadatos* y las relaciones *valorados*, *realizados*, *definidos* y *distribuidos* serán transformadas propagando la clave de la tabla *metadatos* a las tablas *general*, *contacto*, *calidad* y *distribución* con lo que las tablas quedarán como sigue:

metadatos (id, id_estandar, id_version, id_capa, id_idioma)
estandar (id, estandar)
calidad (id, id_metadato, id_fuente, linaje, consistencia, exactitud)
general (id, id_metadato, id_formato, id_sistemaref, fecha_creacion, palabra_clave, descripción, escala)

contacto (id, id_metadatos, id_equipo, id_institucion, id_rol, email, telefono)

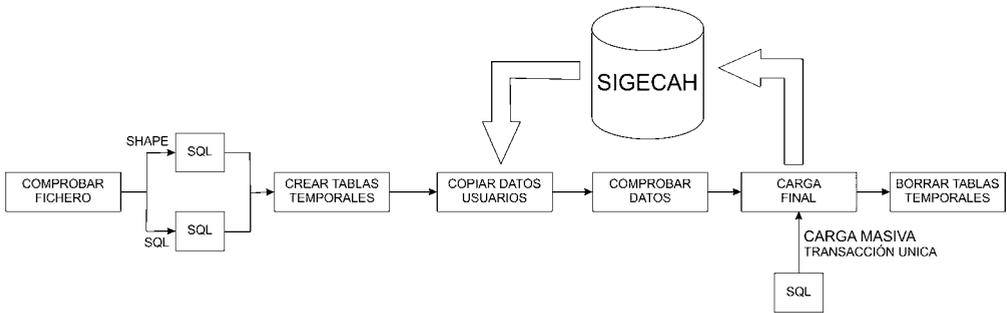
distribucion (id, id_metadatos, id_equipo, id_institucion, id_rol, id_formato_distribucion, email, telefono, restricciones_uso)

5. Implementación en el sistema gestor de base de datos

Todas las tablas del modelo están pensadas para una implementación física directa con dos excepciones: *elemento_geografico* y *elemento_ensenada*.

- Elemento_geografico*. Debido a la forma en la que se producen e introducen los datos en el sistema, se ha optado por almacenar todas y cada una de las capas generadas de manera independiente, lo que ofrece una mayor integración con PostGIS y permitirá una mayor agilidad en la utilización de los datos, evitando así problemas en la actualización. *Elemento_geografico* se implementará mediante una vista que recopilará todos los elementos contenidos en cada una de las capas en un único contenedor o tabla con un identificador único. De este modo será posible identificar unívocamente cada uno de los elementos espaciales (punto, línea o polígono) de cualquier capa.
- Elemento_ensenada*. Esta tabla se implementa físicamente de forma similar a la anterior mediante una vista que almacena el identificador único del elemento de Ensenada a georeferenciar y el tipo de elemento (bien, ubicación, localidad, provincia, etc...).

Figura 6



5.1. Entrada de datos

La entrada de datos puede producirse por dos vías: *shape* o *sql*. Los datos que entren por la primera serán tratados de manera igualitaria tras la transformación del *shape* en un fichero con sintaxis *SQL*. Los datos introducidos vía formato *shape*¹¹ son subidos mediante un mecanismo tipo formulario desde el navegador donde llegan al sistema de ficheros local del servidor. Este sistema permite subir gran cantidad de datos mediante operaciones GET /PUT desde un navegador. Asimismo existen módulos de interfaz de usuario en tecnología FLASH y JavaScript que permiten incluso superar el modelo actual gestionando cargas de ficheros de inmenso tamaño mediante técnicas AJAX.

Una vez leído el fichero, se transforma la lista de vértices en un formato textual de órdenes *SQL* que puede aplicarse en forma de lote, contra la BBDD de esquemas GIS que alberga el portal SIGECAH. Por motivos de seguridad y fiabilidad, la importación se realiza a un esquema temporal donde se pueden aplicar verificaciones adicionales y restricciones a los datos

¹¹ El formato *shape* es propietario, pero existen diversas bibliotecas de libre acceso que permiten leer y verificar dicho formato.

de manera que cumplan tanto las marcadas por las relaciones entre los datos como las derivadas del uso de un sistema multiusuario.

Para facilitar la modificación por lotes se puede exportar los trabajos y análisis ya ejecutados y asimismo volver a importar dichos datos en la base de SIGECAH. Para este propósito el formato *SQL* en el que se exportan los datos se puede volver a cargar de forma directa al esquema temporal donde se reintegra al sistema.

Las bases de datos que contiene SIGECAH tienen potentes herramientas para la verificación de los datos y sus relaciones en forma tabular que se usan en forma de procedimiento almacenados, bien mediante llamadas explícitas desde el portal SIGECAH bien iniciadas bajo demanda en forma de triggers PL/PGSQL (figura 6).

6. Conclusiones

Con este trabajo cerramos la presentación de la *base de datos SIGECAH* iniciado en el número 72 de esta misma revista. A modo de recopilación indicaremos que la base de datos SIGECAH se divide en dos componentes: *Ensenada* y *GIS*. El primero de ellos se corresponde con la fase de alma-

cenamiento de la información textual contenida en el Catastro de Ensenada, mientras que el segundo se corresponde con la extensión espacial del anterior, siendo el encargado de geolocalizar los elementos almacenados. La base de datos tiene como objetivo permitir volcar de manera escalonada e independiente la información textual de la fuente e ir asociada de una cartografía predictiva. No es un modelo cerrado, y se ha diseñado permitiendo la escalabilidad. Actualmente se centra en una parte del corpus documental de la documentación ensenadista, pero en cualquier momento puede ser ampliado para analizar la de cualquier otra fuente del mismo tipo, si existe demanda para ello.

La base de datos descrita se contextualiza dentro del proyecto SIGECAH (Sistema de Gestión de Catastros Históricos), cuya finalidad es facilitar a la comunidad investigadora una herramienta para el procesamiento y análisis de la información contenida en fuentes geohistóricas de carácter textual. Para lograr este objetivo, el proyecto cuenta con un portal web que ofrece al investigador un lugar en el que, sin necesidad de contar con licencias de software y sin conocimientos técnicos, puede encontrar una herramienta que normalice el trabajo con una fuente que contiene una información ingente. En este proceso, y para el caso que nos ocupa (entrada de datos y geolocalización), como ya indicamos en el artículo anterior, contamos con una serie de formularios que permiten interactuar con el modelo presentado en ambas publicaciones.

Bibliografía

- BOSQUE SENDRA, J. (1997): *Sistemas de Información Geográfica*. Madrid, Ed. Rialp.
- CAMARERO BULLÓN, C. (1998): «La cartografía del Catastro de Ensenada, 1750-1756», en *Estudios Geográficos*, 231, págs. 245-283.
- (2002): «Averiguarlo todo de todos: el Catastro de Ensenada». *Estudios Geográficos*, 248-239, págs. 493-531.
- (2002): «El Catastro de Ensenada, 1749-1759: diez años de intenso trabajo y 80.000 volúmenes manuscritos», en *CT Catastro*, 46, págs. 61-88 (español) y págs. 141-153 (inglés) [disponible en red: http://www.catastro.meh.es/esp/ct_catastro.asp] [accedido en 1 de julio de 2012].
- (2007): «La cartografía de los catastros españoles del siglo XVIII», en VV.AA.: *La cartografía cadastral a Espanya (segles XVIII-XX)*. Barcelona, Institut Cartogràfic de Catalunya, págs. 39-52.
- FERNÁNDEZ-WYTTENBACH, A.; SILBATO, W.; MOYA, I.; DAWOOD, Z.; GALINDO A., y BERNABÉ-POVEDA, M. A. (2009): «Integración de Cartotecas Virtuales como herramienta de apoyo en la investigación histórica y social», en *Actas de las Jornadas de la Infraestructura de Datos Espaciales de España (JIDEE 2009)*. Murcia, 4-6 de noviembre de 2009.
- FERRER RODRÍGUEZ, A. (2002): «La documentación del Catastro de Ensenada y su empleo en la reconstrucción cartográfica», en *CT Catastro*, 46, págs. 99-110 (español) y págs. 161-166 (inglés) [disponible en red: http://www.catastro.meh.es/esp/ct_catastro.asp] [accedido en 1 de julio de 2012].
- GARCÍA JUAN, L.; ÁLVAREZ MIGUEL, A. J.; CAMARERO BULLÓN, C., y ESCALONA, J. (2011): «Modelo de datos para la digitalización y gestión de fuentes catastrales geohistóricas textuales. Aplicación al Catastro de Ensenada», en *CT Catastro*, 72, págs. 73-97 [disponible en red: http://www.catastro.meh.es/esp/ct_catastro.asp] [accedido en 1 de julio de 2012].
- GARCÍA JUAN, L.; CAMARERO BULLÓN, C.; FERRER RODRÍGUEZ, A.; TRONU MONTANÉ, C., y ESCALONA, J. (2010): «Propuesta metodológica para la reconstrucción de parcelarios antiguos a partir de fuentes catastrales textuales», en *Actas del XII Coloquio Ibérico de Geografía*, Oporto (en prensa).
- GARCÍA JUAN, L.; ESCALONA, J., y CAMARERO BULLÓN, C. (2008): «Propuesta metodológica para la reconstrucción del parcelario antiguo mediante sistemas de información geográfica», en *CT Catastro*, 63, págs. 203-214 [disponi-

- ble en red: http://www.catastro.meh.es/esp/ct_catastro.asp] [accedido en 1 de julio de 2012].
- GRAU MIRA, I. (ed.): (2006): *La aplicación de los SIG en la arqueología del paisaje*. Publicaciones de la Universidad de Alicante, Alicante.
- HEINEMAN, G. y COUNCILL, W. (2001): *Component-Based Software Engineering: Putting the Pieces Together*. Addison-Wesley Professional, Reading.
- VV.AA. (2007): *La cartografía cadastral a España (siglos XVIII-XX)*. Barcelona, Institut Cartogràfic de Catalunya.